REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 28-03-2018	**2. REPORT TYPE** research	**3. DATES COVERED** *(From - To)* October 2017-March 2018

4. TITLE AND SUBTITLE

Cybersecurity Assessment Parameter Profile (CAPP)
A Tool for Making Sense of Cybersecurity Assessments

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Russell, Scott, L
Jackson, Steven, C

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Surface Warfare Center, Crane Division
300 Hwy 361
Crane, IN 47522

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Naval Surface Warfare Center, Crane Division
300 Hwy 361
Crane, IN 47522

10. SPONSOR/MONITOR'S ACRONYM(S)

NSWC

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This paper seeks to answer three core questions: How can decision makers (1) identify the salient differences between existing cybersecurity assessments; (2) select the most appropriate cybersecurity assessments for their missions, resources, and constraints; and (3) find and fill gaps in the cybersecurity assessment ecosystem? The framework uses a cohesive set of eight parameters to characterize cybersecurity assessments, and introduces the Cybersecurity Assessment Parameters Profile (CAPP) tool, which aids decision makers in applying the parameters to cybersecurity assessments. Each parameter is a non-categorical spectrum, whose extremes offer both utility and limitations. Each parameter offers a meaningful choice for cybersecurity decision makers, as every parameter value is desirable for some assessment scenario.

15. SUBJECT TERMS

Cybersecurity

16. SECURITY CLASSIFICATION OF: U			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Robert Templeman
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UU	33	**19b. TELEPHONE NUMBER** *(include area code)* 812-854-6565

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Cybersecurity Assessment Parameter Profile (CAPP)

A Tool for Making Sense of Cybersecurity Assessments

March 2018

Scott Russell and Craig Jackson

Naval Surface Warfare Center, Crane Division

Acknowledgements
The authors thank Dr. Robert Templeman and Mr. Kelly Richmond for their support and feedback, as well as Indiana University colleagues Dr. Dan Hickey, Mr. Mark Krenz, Mr. Anurag Shankar, and Mr. Ryan Kiser for their feedback on concepts included in this work.

Funding
This work was funded under the Naval Innovative Science and Engineering (NISE) program (Sec 219).

Audience for this work
The audience for this work is any decision maker with a practical need to understand the variety, pros, and cons of real world cybersecurity assessments. These decision makers may be responsible for selecting, commissioning, designing, facilitating, conducting, or reviewing cybersecurity assessments, evaluations, and audits. We assume that the reader has some familiarity with cybersecurity concepts, but may not have a great deal of specialized expertise.

Table of Contents

Executive Summary

This deliverable seeks to answer three core questions: How can decision makers (1) identify the salient differences between existing cybersecurity assessments; (2) select the most appropriate cybersecurity assessments for their missions, resources, and constraints; and (3) find and fill gaps in the cybersecurity assessment ecosystem?

The framework presented here uses a cohesive set of eight parameters to characterize cybersecurity assessments, and introduces the Cybersecurity Assessment Parameters Profile (CAPP) tool, which aids decision makers in applying the parameters to cybersecurity assessments. Each parameter is a non-categorical spectrum, whose extremes offer both utility and limitations. Each parameter offers a meaningful choice for cybersecurity decision makers, as every parameter value is desirable for some assessment scenario.

The three primary use cases for the Parameters and the CAPP tool are:

1. <u>Understanding and selecting among competing cybersecurity assessments</u>: The parameters provide a single, comprehensive framework for understanding what a particular assessment's utility and limitations are, and which assessment(s) are best suited to address the needs of a given mission.
2. <u>Building a portfolio of cybersecurity assessments</u>: Organizations frequently must leverage a range of cybersecurity assessments to meet all their needs (e.g., over a system's acquisition and operation lifecycle). The parameters and profiles can be used flesh out a portfolio of assessments, and avoid unnecessary redundancy.
3. <u>Structuring conversations between stakeholders and assessors</u>: Finally, the parameters provide a basic structure for conversations among stakeholders and assessors. Both can use the parameters to clarify what the assessor's assessment provides and what the stakeholder actually needs.

The Parameters:

Substantive Parameters. Substantive parameters relate to what the stakeholder wants from the assessment. Substantive parameters describe the scope, output, and standards of the assessment.

1. Advisory and Informational Deliverables. *Will the assessment deliverables focus on providing information, advice, or both?*
2. Broad and Narrow Assessment Targets. *How is the assessment target scoped?*
3. Minimum, Maximum, and Tailored Standards. *What type of standard will the target be measured against?*

Procedural Parameters. Procedural parameters relate to how the assessment will be conducted. Procedural parameters describe what the assessor performs, including their methodology, tools, and techniques.

1. Standardized and Specialized Methodologies. *Is this assessment standardized, or dependent on specific assessors?*
2. Live, Conceptual, and Modeled/Simulated Settings. *Does the assessment engage with live systems, models and simulations, or only abstracted scenarios?*
3. Internal and External Assessors. *Who is conducting the assessment: me, someone in my organization, or a third party?*
4. Red and Blue Assessor Roles. *Is the assessor taking the perspective of an attacker or a defender?*
5. Quantitative and Qualitative Analyses. *Is the assessment focused on quantified data or qualitative observations and analyses?*

1. Introduction: Making Sense of Cybersecurity Assessments

Cybersecurity assessments, at their core, are a means for decision makers to improve their informational posture in order to better confront challenging cyber scenarios. They can provide decision makers a clearer understanding of their own organization, their enemies, their environment, and a host of other factors that may be relevant to cyber-decisionmaking. As such, cybersecurity assessments are a diverse and complex topic, offering a range of perspectives and deliverables, and the selection of a specific cyber-assessments or set of assessments will derive from the specific needs of the decision maker. **But in all cases, the goal of the cybersecurity assessment is to inform cybersecurity decisionmaking.**

As cybersecurity has risen in prominence, so too have the number and variety of cybersecurity assessments, and the number of entities developing, implementing, and commissioning those assessments. We observe risk assessments, threat assessments, compliance audits, red team exercises, blue team assessments, table top exercises, resilience reviews; all of which are supposed to assess *something* relevant to cybersecurity and impact future decisionmaking. Adding to this confusion, cybersecurity assessments and assessors can be unclear as to what exactly they provide: different actors use these terms to mean different things; subjective assessments may offer false "quantitative" weight; narrow quantitative assessments may be interpreted too broadly; and technical compliance checklists may be touted as holistic programmatic assessments. Like so many other areas of cybersecurity practice, the realm of cybersecurity assessments is relatively immature. There is a great deal of experimentation and capitalization, and a fair amount of confusion.

At the same time, there is an increasing appreciation that cybersecurity assessments are necessary and valuable tools to influence and inform cyber-decisionmaking on where to move precious resources, when to deploy new controls, and when enough is enough. In this context, the decision makers who must grapple with cybersecurity assessments may be unclear about exactly what assessment they need or want. More fundamentally, they may be unclear as to what cybersecurity assessments actually do.

<u>Goal</u>. Our goal with this paper is to answer three core questions: How can decision makers: (1) Identify the salient differences between existing cybersecurity assessments; (2) Select the most appropriate cybersecurity assessments for their missions, resources, and constraints; and (3) Find and fill gaps in the cybersecurity assessment ecosystem?

<u>Roadmap</u>. In Section 2 we provide context for our research, framing the work with our operational definition of "cybersecurity assessment." In Section 3, we identify and delineate the basic parameters that our research uncovered, and provide insight into when and why the values on those parameters can prove useful.[1] In Section 4, we apply the parameters to sample cybersecurity assessments. In Section 5 we conclude with a set of use cases for the parameters.

2. Background: What is a "Cybersecurity Assessment"?

The US Navy defines cybersecurity as "the prevention of damage to, protection of, and restoration of computers, electronic communications systems, electronic communications services, wire communication,

[1] We include a complete discussion of our methodology in Appendix B.

and electronic communication, including information contained therein, to ensure its availability, integrity, authentication, confidentiality, and nonrepudiation."[2]

We were unable to find a single, widely agreed-upon definition of **"cybersecurity assessment"** in the literature.[3] Even the as-used definition of risk assessment has required substantial unpacking.[4] Moreover, our research confirms that there exists a very broad range of assessments designed to inform cybersecurity-related decisionmaking. For the purposes of this paper, our operational definition of "cybersecurity assessment" is as follows:

> A cybersecurity assessment is an analytical activity directed at an identified target, whose outputs' purpose is to inform stakeholder cybersecurity decisionmaking regarding: (a) the characteristics and appropriate operational roles of the target; (b) the target's readiness to operate; and/or (c) resources, policy, processes, and controls warranted to support the target's operational use.[5]

Within this definition, two additional terms require clarification, as we use them throughout the paper. We use this term **"target"** to refer to the scope of the thing the assessment will assess.[6] Cybersecurity assessments can be directed toward a great variety of targets, including organizational cybersecurity programs, plans and activities of specific missions, platforms, systems of systems, systems, and system components (including hardware, software, or human elements).

The second term is **"stakeholder."** A stakeholder is an entity in a position of authority, ownership, control, or operation vis-à-vis the target. In our discussions, the stakeholder is typically the person or entity for whom the assessment deliverable is being produced. Stakeholders include entities and individuals who are organizationally or operationally "close" to the target (e.g., the owner, operator, or developer of the target system), and those relatively distant (e.g., an oversight group or strategic leadership).

3. Common Cybersecurity Assessment Parameters

Using an iterative, top-down and bottom-up research approach,[7] we identified eight common cybersecurity assessment parameters broken into two categories: Substantive parameters and Procedural parameters. These

---

[2] National Security Presidential Directive 54/Homeland Security Policy 23, "Cybersecurity Policy," White House, 8 Jan. 2008, https://fas.org/irp/offdocs/nspd/nspd-54.pdf.

[3] For instance, although NIST has published a definition of "assessment" in its Glossary of Key Information Security Terms, this definition only points to the narrower "Security Controls Assessment," which is defined as "the testing and/or evaluation of the management, operational, and technical controls in an information system to determine the extent to which the controls are implemented correctly, operating as intended, and producing the desired output with respect to meeting the security requirements for the system and/or enterprise." We define "cybersecurity assessment" more broadly than this.

[4] *See, e.g.,* Nachtigal *et al., "*Analysis of Alternatives for Risk Assessment Methodologies and Tools," Sandia, Oct. 2013, http://prod.sandia.gov/techlib/access-control.cgi/2013/138616.pdf.

[5] In Appendix A, we provide a grounding analysis for each element of the definition.

[6] We also discuss these differing assessment targets as a distinct parameter in Section 3.1.2.

[7] For a complete discussion of our methodology, see Appendix B.

two categories all relate to different aspects of the assessment. Substantive parameters are those primarily focused on the purpose and outputs of the assessment, whereas Procedural parameters are those focused on how and by whom the assessment is conducted.

Within these categories, a core requirement for each parameter is that every potential value must be useful and desirable under some circumstances. For instance, on the Red/Blue parameter, both "red" and "blue" assessments have scenarios where a stakeholder would actively choose that value over the other. Both have unique utility and limitations, and will be better suited for different scenarios. We distinguish this from "always good" parameters, where only one value is desirable, such as the competence of the assessor, or the clarity of the deliverable.[8] (In no scenario is a "low competence" assessor preferable.) Here, we focus on areas which present stakeholders with a *meaningful choice*, and so we have not included parameters where there is a clear best answer in all circumstances.

The remainder of this section is broken into three parts, based on the three categories of parameters we identified: Section 3.1, Substantive Parameters**;** and Section 3.2, Procedural Parameters.

3.1 Substantive Parameters *("What do you want from the assessment?")*

The first category of parameters is those that relate to the output of the assessment. What are the substantive requirements that determine what the assessment will generate? These substantive parameters govern what the assessment does: Does the assessment produce an informational report, or does it offer pointed advice on what to do, and why? Does the assessment's scope only cover a single critical system, or is it assessing an entire command? What type of standard will the target be assessed against?

Substantive parameters are most closely tied to what the stakeholder wants to get out of the assessment, and will go on to implicate how the assessment should be conducted.

3.1.1 Advisory and Informational Deliverables

The advisory/informational spectrum is important for selecting where the focus of the assessment will lie. Is the focus of the assessment to provide information as comprehensively and dispassionately as possible, or is the focus in crafting recommendations that will maximize the benefit to the assessment target? Or does the assessment attempt to balance the two? It is critical when selecting a cybersecurity assessment to consider whether you are only seeking factual information, if you need specific guidance to help parse that information, or whether you really just need to be told what to do.

Advisory deliverables are those where the deliverable explicitly provides the stakeholder with advice, recommendations, action items, or priorities. These assessments seek to guide decisionmaking directly, and therefore inject the opinions and expertise of the assessor into the stakeholders' decisionmaking processes.

[8] However, an important subset of these "always good" parameters is "trade-offs." Trade-offs are competing interests that stakeholders must balance when selecting an assessment (e.g., cost, schedule, risk, reliability).[8] Although each individual trade-off has a clearly preferable value, that value cannot be selected without negatively impacting other trade-offs. For instance, a low-cost assessment is always preferable to a high-cost one, but selecting a low-cost assessment negatively impacts other trade-offs, such as the reliability of the deliverable. Identifying and managing these trade-offs is an important preliminary step when selecting between cybersecurity assessments.

Advisory assessments may take the form of a binary "yes/no" determination about whether a system should be authorized to operate, may offer a prioritized list of recommendations, or may offer strategic guidance on how to improve the target's security strategy.

> **Utility and limitations:** Advisory deliverables' primary utility is their ability to provide actionable, practical guidance. Advisory deliverables directly leverage the expertise of the assessor or otherwise direct action. This makes advisory deliverables more likely to be immediately useful to stakeholders, as they provide actionable advice on how to proceed. But this also means that advisory assessments are highly assessor-dependent, and their strength or weakness can hinge on the specific assessors. This can be a source of both utility and limitations, depending on the assessor. Moreover, the quality of advisory deliverables hinges of the assessor's access and ability to understand the stakeholder's goals, priorities, capabilities, and risk appetite. And at the extreme end of the spectrum, advisory deliverables without informational grounding can be very difficult to assess for factual relevance. Finally, advisory assessments are more susceptible to assessor biases, particularly when the assessor has a financial interest in some of the recommendations in the deliverable.

Informational deliverables offer factual information to the stakeholder. Deliverables with an informational focus prioritize the dispassionate communication of factual information. The goal of these deliverables is to *inform* stakeholder decisionmaking, and leave judgment and action-taking to the ultimate decision makers. Informational deliverables are highly variable in their subject matter, ranging from catalogues of security controls to information about the threat environment. Indeed one of the most familiar cybersecurity assessments is a "risk assessment," which typically is purely informational. Informational deliverables, low on advisory content, may be particularly useful for organizations with mature cybersecurity programs, as those organizations will have the expertise and experience to comprehend and apply the information they receive into their cybersecurity decisionmaking processes.

> **Utility and limitations:** Informational deliverables' primary strength is their factual objectivity and verifiability. By focusing significant attention on the collection and presentation of information, informational assessments have less potential for hidden bias, are more easily evaluated, and are more easily replicated by other assessors. They focus the limited resources of time, effort, and money on building a factual picture. In many cases this dispassionate factual picture is preferable for decisionmaking. However, when focusing on *just the facts*, informational deliverables require the decision maker to understand the context for those facts, evaluate the weight and merits of those facts, and craft the response. Fundamentally, informational assessments are not by themselves actionable: they only serve to inform someone else's decisionmaking process. This makes them ripe for misinterpretation, over-extension, misapplication, or the collection of dust.

3.1.2 Broad and Narrow Assessment Targets

The second deliverable parameter is the scope of the assessment. Put simply, what is being assessed? Is it a mission, a platform, a system of systems, a single system, a system design, subsystems, components? Is it assessing a target in isolation, or in the context of its users and operational environment? Assessments can range from high-level reviews of the cybersecurity program for an entire mission, command, or platform, to deep-dive assessments of critical systems, subsystems, or components. Some assessments may dynamically balance the two extremes, beginning with a broad analysis, and then selectively employing greater depth for

certain critical paths. Cybersecurity concerns can arise at all levels of organizational complexity, so the selection of a given assessment must take into account what specifically is being assessed, and how much broader context that assessment should incorporate.

Broad target definition scopes assessments to complex systems, systems of system, platforms, commands, or other multifaceted activities. Rather than focus on specific minutiae, broad assessment targets support a more holistic perspective, focused on comprehensive coverage, and importance to the mission. Broad assessment targets focus on complex systems, and place particular emphasis on the high-level interactions that those complex systems engage in. Finally, broad assessment targets can serve as a foundation for more narrowly scoped efforts in the future.

> **Utility and limitations:** Broad assessments' primary strengths are their comprehensive, holistic perspectives. Broad scoping facilitates inclusive assessments that better account for mission assurance, critical gaps in programmatic security, and the broader context in which technologies and processes must operate. By defining the assessment target broadly, stakeholders make sure they don't lose sight of the proverbial forest for the trees. Broad assessments also benefit from more wholesale applicability, as the information they provide will prove useful across the stakeholder's environment, as opposed to selectively improving a few key spots. The limitations of broad target definition follow naturally from its strengths: In the context of limited resources, breadth limits depth. Broad assessments cannot offer the level of depth, detail, and complexity that a narrow assessment can. Broad assessments can also prove difficult to execute in practice, as their broad scope requires understanding a wide range of organizational and managerial concerns, and may necessitate a more diverse team of assessors to tackle the wide-ranging issues.

Narrow target definition scopes assessments to specific systems, components, or problem areas. Narrow scoping prioritizes detail and concentrates resource usage. Narrow scoping aids close looks at mission critical systems, high-traffic nodes, or areas of high vulnerability (like human operators, hastily produced code, or legacy technologies). Narrow assessments are also more likely to engage with technical minutia, and their output will often be targeted to the level of complexity being addressed.

> **Utility and limitations:** The primary utility of narrow scoping is it should focus detailed attention on the most valuable and vulnerable assets. By restricting the scope of the assessment, narrow assessments provide sharper focus and greater detail on the assessed subject matter. Nevertheless, narrow assessments have notable limitations. The tighter focus can make the assessments' significance difficult for higher-level decision makers to understand and contextualize, and the fine-grained nature of the problems they uncover can be difficult to understand in the context of the mission. This complexity also makes narrow assessments potentially challenging for smaller, less mature organizations to take action upon. And more fundamentally, narrow assessments may lack the big picture perspective that comes with broad assessments, increasingly the likelihood that the assessment won't even look at major problem areas because they are out of scope.

3.1.3 Minimum, Maximum, and Tailored Standards

Cybersecurity assessments vary in terms of the type of standard they employ. "Assessments are comparisons,"[9] and the standard is what the target is compared against. Is the assessor comparing against a hypothetical perfect or "maximum security," determining if the security passes bare-minimum requirements, or determining if the security is well-tailored for the target's needs? The standard is the critical piece of context for understanding and interpreting an assessment deliverable. It frames the meaning and import, providing context for the information or recommendations the assessment provide. A particularly onerous recommendation will be interpreted very differently if it is based in a minimum standard as compared to a tailored or maximum standard. Assessments may employ more than one type of standard, so it is particularly important for stakeholders to clarify, "what are we being measured against?"

Minimum standards are those used to compare the target to a set of minimum requirements (e.g., a list of legal obligations to implement a defined set of controls).[10] Minimum standards look only to the baseline requirements, and thus are most commonly associated with compliance. But minimum standards can include any assessment that evaluates against a minimum requirement, including a requirement that an assessor creates. Minimum standard assessments are therefore best thought of as a filter: They identify states of affairs that are adequate and those that are inadequate, but may provide little information beyond separating those who do and do not meet the minimum requirements.

> **Utility and limitations:** The primary strength of utilizing minimum standards is their relative ease of application and understanding. Minimum standards establishing a baseline level of security. Thus, the requirements tend to be straightforward, limited, and of low-complexity. Similarly, understanding a minimum standards assessment is rarely a problem for stakeholders, as the output can be as simple as a Pass/Fail grade, where a failing grade may also include a report highlighting areas for potential improvement. This simplicity and *a priori* determination of minimum standards facilitates a greater degree of objectivity, repeatability, and standardization. The limitations of minimum standards assessments follow naturally: Minimum standards do not assess "security" in a broad sense; they assess whether the target meets minimum requirements. Assessments strictly using minimum standards may not provide any information above this baseline, making them of little use to organizations that have already met the minimum, or for whom the minimum is ineffective or inappropriate.

Maximum standards are those used to compare the target's security to a theoretical maximum or perfect state.[11] Maximum standards attempt to leave no holds barred, and will nitpick every potential vulnerability, inefficiency, or insufficiency in the target. But more fundamentally, maximum standards embrace a mindset of unrelenting improvement, focusing not on satisfying a fixed standard but on identifying what can be done better. This quality makes maximum standards a potential springboard for innovation. When the stakeholder wants to understand their security against the most advanced actors possible, they will seek out maximizing assessments.

---

[9] Campbell, P., & Stamp, J, "A Classification Scheme for Risk Assessment Methods," Sandia, Aug. 2004, *available at* http://energy.sandia.gov/wp-content/gallery/uploads/sand_2004_4233.pdf.

[10] *See, e.g.,* "Payment Card Industry – Data Security Standard v. 3.2," PCI-SSC, April 2016, https://www.pcisecuritystandards.org/document_library.

[11] *See, e.g.,* Crown Jewels Analysis, MITRE Corp., *available at* https://www.mitre.org/publications/systems-engineering-guide/enterprise-engineering/systems-engineering-for-mission-assurance/crown-jewels-analysis.

> **Utility and limitations:** The primary utility of utilizing maximum standards is that they offer the highest standard of by which to assess cybersecurity for the target. When assessors utilize these extreme standards, they put themselves in the position to explore every aspect of "what could go wrong." Assessments against maximum standards may provide stakeholders with the most information. This makes them ideal for targets with very high security needs, or for mature organizations that want to understand their security in absolute terms. However, assessments focused on maximum security are likely to come with a host of information or recommendations that is of little practical use. If an organization does not have the resources or expertise to strive for perfection, then why bother? Assessments that focus on maximum security may not only produce unwieldy amounts of information or unrealistic recommendations, but may consume disproportionate resources for the value they provide and fail to align with stakeholders' needs and the targets place in the mission.

Tailored standards are those compiled to strike an ideal balance of security for the target. Tailored standards are used to evaluate the target against where the target *should* be, rather than against some preexisting standard that was developed without regard for the target's mission. Unlike maximum standards, which evaluate against security in the abstract or against the extreme case, tailored standards evaluate against a target-specific standard that incorporates the target's mission, priorities, and constraints. Tailored standards put security in the context of the target's mission, and can be quite useful for less mature targets that would prefer the assessor determine what is best for their mission.

> **Utility and limitations**: The primary strength of tailored standards is that they evaluate against the target's ideal state, a theoretically optimal standard. Tailored standards, implemented properly, provide an optimized blend of security advice. Tailored standards also most directly impart the expertise of the assessor. Additionally, use of tailored standards provides flexibility, allowing for the nuances of the target's mission to be reflected in the deliverable. Nevertheless, use of tailored standards can have a host of limitations. They can be difficult to evaluate objectively, are highly assessor-dependent, and are the most unreliable standard, as the standard itself is subject to assessor interpretation. Indeed the difficulty in crafting and implementing tailored standards means that they often devolve to minimum standards, maximum standards, or some combination of the two. (We observe this in environments that have attempted to embrace complex risk management frameworks.) Tailored standard are also increasingly difficult to perform on more complex targets, and for targets that do not understand their own mission, priorities, and risk tolerance. And more fundamentally, tailored standards are the most susceptible to assessor biases, as the assessor is not relying on a preexisting standard, but is instead creating a unique standard for each target.

3.2 Procedural Parameters *("How is the assessment conducted?")*

If the substantive parameters cover the *why* and *what* of an assessment, procedural parameters describe *how*, *when*, *where*, and *by whom* the assessment is conducted. Procedural parameters are primarily concerned with the internal workings of the assessment: How was it conducted, who will carry it out, what methods will be utilized? While Procedural parameters are more often the purview of the assessor, savvy stakeholders may still require certain procedural parameters. In all cases it is important for the stakeholder to understand how these different procedural parameters may impact the assessment process, and be prepared to identify the utility

and limitations that specific types of assessments will offer. The goal is to unwind the internal workings of cybersecurity assessments, helping stakeholders understand and articulate what they want, while helping assessors better understand the key distinguishing parameters of their own methodologies.

3.2.1 Standardized and Specialized Methodologies

Standardization refers to the degree to which the assessment follows a clear, repeatable, assessor-neutral methodology, as opposed to a more flexible, closely held, proprietary, or ad hoc methodology.[12] More standardized methodologies are typically associated with high volume, low-complexity assessments, like compliance audits, whereas specialized assessments are typically associated with independent expert analyses. Indeed, some assessments may appear at first glance to be specialized, (say, because only one assessor can conduct it,) but actually adhere to a highly standardized methodology. In such a case, the barrier to other assessors conducting the same assessment is not fundamental to the assessment itself, but rather the lack of documentation explaining how. While in practice most assessments will involve some degree of both standardization and specialization, this parameter is primarily concerned with where the focus lies: Is the assessment predominantly standardized, predominantly specialized, or somewhere in the middle?

Standardized methodologies are transparent (e.g., documented), rigid, repeatable, and assessor-neutral.[13] They are ideal for higher volume, lower-complexity assessment targets, for benchmarking, and are particularly useful for pairing with quantified analyses. However, standardized assessments can also be used in qualitative assessments, such as with the Delphi Method,[14] as the core of standardization is that there is a clear structure guiding how the assessment methodology is conducted. A standardized assessment will be laid out such that any competent independent assessor can pick it up and perform it effectively.

> **Utility and limitations**: Standardization makes assessments more repeatable, more easily assessed, evaluated, and improved, and more appropriate for competitive environments (e.g., where a winning technology must be selected). Standardization enables scaling the assessment to more assessors and more targets. Standardization also helps mitigate the impact of human biases, conflicts of interest, and isolated failures, by relying on the structure of the assessment, rather than individuals, to provide value to the stakeholders. The limitations of standardized assessments arise naturally from this rigidity. Standardized assessments are necessarily inflexible, narrow, and restrictive. Standardized methodologies are fundamentally limited in what they can assess, as the subject matter must be amenable to algorithmic analysis and *ex ante* evaluation. This also means that standardized assessments are drawn toward data points that are more easily evaluated, regardless of their salience to the broader mission. And perhaps most troublingly for cybersecurity, standardized assessments are poorly suited to address rapidly changing problem areas, as each shift requires a potential reevaluation of the entire assessment methodology.

---

[12] Note, whether the *methodology* is standardized has no bearing on what type of standard the target is being assessed against. A standardized methodology can be used to assess a target against a minimum, maximum, or tailored cybersecurity standard.

[13] *See, e.g.,* "Cybersecurity Test and Evaluation Guidebook," Department of Defense, 1 July 2015, *available at* http://www.dote.osd.mil/docs/TempGuide3/Cybersecurity_TE_Guidebook_July1_2015_v1_0.pdf.

[14] The Delphi Method is product of the RAND Corporation, and provides a structured approach for the solicitation of expert opinion that accounts for traditional group decisionmaking biases. *See, e.g.,* Delphi Method, RAND, https://www.rand.org/topics/delphi-method.html.

Specialized methodologies, by contrast, are those that forgo strict procedural rigor in favor of greater flexibility and more discretion on the part of the assessor. These assessments are designed to tackle more varied and diverse subject matter, and recognize that strict methodologies may be inappropriate when confronted with unconventional problems, nontraditional missions, or limited resources. By specializing, these methods empower the assessor to innovate, deviate, and otherwise make strategic decisions that greater standardization would not allow.

> **Utility and limitations**: The primary strength of specialized methodologies is their flexibility, allowing for greater fine-tuning with regard to deliverables, priorities, and costs. This greater flexibility of specialized methodologies makes them better suited for addressing rapidly changing problem areas. Specialized methodologies also benefit from more effectively harnessing the expertise of their assessors, something more rigid standardization can hinder. The limitations of specialized methodologies follow logically. Specialized methodologies scale poorly, and can be difficult for potential customers to access. Specialization also makes assessments harder to evaluate, both *ex post* and *ex ante*; specialized methodologies are difficult to reproduce, study, and compare with other assessments; and specialized assessment methodologies' greater flexibility can allow for bad practices to go unnoticed and unremedied. Specialized methodologies often rely on the expertise of a select few assessors, and may have little underlying structure other than the routine practices of the assessors at issue. This also means that specialized assessments can rely more heavily on the skill of the assessor. These limitations raise concerns about the reliability, veracity, and weight that can be accorded more specialized assessments.

3.2.2 Live, Conceptual, and Modeled/Simulated Settings

The second procedural parameter represents how closely the assessor engages with the assessment target. The live/conceptual/modeled/simulated parameter is about what the assessor actually looks at: are they looking at the target operating in real time? Are they looking at models and simulations of the target? Or are they looking at documentation about the target? Although each of these operates on a spectrum, the fundamental question is "how close is the assessment to reality?" Although "proximity to reality" can vary on a number of fronts, from the environments the target operates in to the threats the target may face, we have chosen to particularly emphasize the proximity to the target itself.

Live assessment settings are those where the assessor directly interacts with the target, watching it operate in real time and subjecting it to stresses to see how it responds. Conceptual assessment settings, by contrast, operate at arm's length, reviewing documentation, policy, procedure, and interviewees, without directly engaging with the live target. Modeled/Simulated assessment settings seek a middle ground, using models and simulations to create a facsimile of the target that approximates real world environments. Although frequently paired to red and blue assessor roles, the red/blue parameter is primarily concerned with the mindset and role of the assessor, whereas the live/conceptual/modeled/simulated parameter is primarily concerned with how closely the assessment setting reflects reality.

Live assessment settings are those that engage the target system in real[15] or near-real world scenarios to evaluate how the target responds to failures, crises, or cyber-decisionmaking scenarios in practice.[16] Live assessments evaluate the system with up-close inspection, often under stress. Live assessment settings can range from live penetration tests to security exercises evaluating policy implementation to site-inspections that determine if configurations match specification. Live assessments are where assessors get to see how the target operates in motion, and identify failures that emerge from complex interactions. Note, additionally, that a retrospective review of a live security incident would still be considered a "live" assessment, as the information informing the assessment was generated by live settings.

> **Utility and limitations:** The primary strength of live assessment settings is that they offer a much closer analogue to real world scenarios. Live settings allow the assessor to quickly identify critical vulnerabilities that would have otherwise stayed hidden in the complexity of system documentation. After all, emergent properties of a complex system are often not discovered until the system is realized, particularly if reality doesn't match documentation. Live settings are therefore particularly useful in high complexity systems or those where documentation is sparse or inaccurate. Live settings are also particularly valuable to help familiarize decision makers and stakeholders with their own processes for incident response. The primary limitation of live settings is their potential for disruption and damage. Live assessment settings are the most organizationally disruptive to conduct, as they entail the assessors interacting with the target *in situ*. Live settings also can incur damage if the assessment includes more adversarial techniques. This makes live settings technically and logistically complicated to operate. Furthermore, the results from live settings are typically narrow in scope, as live assessment settings are particularly challenging at large scales.

Conceptual settings are those where the assessment focuses on the conceptual operation of the target, and evaluates this security in the abstract.[17] Conceptual assessment settings focus on how the target systems *should* operate, and therefore largely avoid analysis of bugs, policy enforcement, or other more practical security failings. Conceptual assessments rely primarily on documentation, interviews/interrogatories, and other arms-length interactions with the target to form their information base, and will take the target at their word that the system operates as presented to them. Although conceptual settings are naturally well-suited for systems in the conceptual phase, the methodology underlying conceptual settings assessments can be useful at any point in the life cycle.

> **Utility and limitations:** The primary strength of conceptual settings assessments is that they are able to abstract away from common, low-level failures, and evaluate the system's security under the best circumstances. Rather than get bogged down with fixing bugs, conceptual settings allow the assessors to focus their energy on larger, more fundamental problems. Conceptual settings are also frequently less resource-intensive, as the fact-finding phase is much more limited, without the need for in-person verification or tabletop exercises. The limitations of conceptual settings arise naturally

[15] Note, the most extreme "Live" assessment setting would be analysis of the target while under attack by an actual adversary. However, for practical reasons, these are relatively rare.
[16] *See, e.g.,* Northcutt et al., "Penetration Testing," SANS Institute, Feb. 2018, *available at* https://www.sans.org/reading-room/whitepapers/analyst/penetration-testing-assessing-security-attackers-34635; "Controls Attacks and Kinetic Effects," NSWC Philadelphia.
[17] *See, e.g.,* "Security Architecture and Design Assessments," Secureworks, https://www.secureworks.com/capabilities/security-risk-consulting/security-design-architecture-assessments; "Cyber Security Architecture Assessment," TBG Security, https://tbgsecurity.com/cybersecurity-architecture-assessment/.

from their strengths. Conceptual settings are fundamentally more detached from reality, and can easily miss major problems that in-person inspection would readily identify. Conceptual settings are also only able to evaluate what the target has documented in some form, and the sophistication of the analysis is limited by the detail the target provides. And perhaps most importantly, conceptual settings miss out on the value that comes from watching complex systems operate under stress, where most failures become visible.

Modeled/Simulated settings are those that use models and simulations to create or use a facsimile of the target to evaluate how that facsimile performs under different realistic scenarios.[18] Modeled/Simulated assessment settings attempt to strike the balance between live and conceptual settings by creating a mock live scenario where the stakes are much lower. Modeled/Simulated assessment settings use tools like computer models or other stand-ins for live settings, and evaluate how that simulated setting performs.

Utility and limitations: The chief utility of modeled/simulated settings is that they allow for the assessment of the target in near-live scenarios, identifying failures that would typically require live settings without incurring the disruption and risk that live settings usually entail. Modeled/Simulated settings offer a middle ground between the detachment of conceptual settings and the risk of live ones. Modeled/Simulated settings also allow for greater use of computation and optimization, as many simulated settings can be run numerous times, to experiment with what has the greatest impact, or what scenarios carry the greatest risk. This potential for repetitive analysis and review is notably different from both live and conceptual settings, which typically cannot be repeated without significant duplication of labor. But the limitations of modeled/simulated settings are just as pronounced. Chief among these limitations is that modeled/simulated settings are not live settings, and their accuracy hinges upon the validity of the model or simulation being used. An unrepresentative model or simulation is at best of low value, and at worst actively misleading. This latter point goes to the more fundamental problem of confidence: Modeled/Simulated settings are fundamentally less reliable than live ones, and modeled/simulated settings also constrain the role of the assessor, removing the primary source of confidence in conceptual settings.

3.2.3 Internal and External Assessors

Assessments can range from self-assessments conducted by target's organizational owner or the owner's agent,[19] to assessments conducted by internal, but organizationally distinct units (e.g., "internal audit"),[20] to external assessments conducted by third parties (e.g., under contractual relationship). As with all the substantive and procedural parameters, the extremes on the spectrum have both utility and limitations that stakeholders need to understand. Many organizations use both internal and external assessors, as well as those

[18] *See, e.g.,* Garvey, P. & Pinto, C., "Introduction to Functional Dependency Network Analysis," MITRE Corp., 2009, *available at* https://pdfs.semanticscholar.org/865c/27f6870ead4fddc7ab0af3248f89f1875dc7.pdf.

[19] For examples of cybersecurity self-assessments, *see, e.g.,* NCCIC Cyber Assessment Tool, ICS-CERT, *available at* https://ics-cert.us-cert.gov/sites/default/files/FactSheets/ICS-CERT_FactSheet_CSET_S508C.pdf; FFIEC Cybersecurity Assessment Tool, FFIEC, *available at* https://www.ffiec.gov/cyberassessmenttool.htm; or the Baldrige Cybersecurity Excellence Builder, *available at* https://www.nist.gov/sites/default/files/documents/2016/09/15/baldrige-cybersecurity-excellence-builder-draft-09.2016.pdf.

[20] *See, e.g.,* Mazmanian, A., "Whitehouse renews call for cyber IG," FCW, 1 Nov. 2017, *available at* https://fcw.com/articles/2017/11/01/senator-whitehouse-cyber-ig.aspx (discussing creation of a cyber investigator general).

positioned to straddle the distinction, in order to reap the benefits and mitigate the limitations of focusing on one extreme.

Internal assessors are those that are organizationally close to the target and its owners and operators.[21] Internal assessments may vary from the project lead self-assessing their own progress, to internal assessment organizations that are organizationally distinct (for example, the office of Inspector General (IG) in the Federal Government).[22] In all cases, the salient factor is that internal assessments are *internal;* they do not seek outside help or advice, and limit the assessment to entities within their larger organizational structure.

> **Utility and limitations**: Internal assessors may benefit from intimate familiarity with the target and its place in the organization's mission, and low overhead[23] during the assessment process. Internal assessments offer the target a structured way to review their own security without negotiating with and putting trust in an outside entity. They may be cost-saving. They are particularly useful when conducted at a higher frequency as a means of tracking improvements over time and in preparation for external assessments. However, internal assessments may be lower sophistication, are more susceptible to self-serving biases, and are more likely to overlook problems that a third party with fresh eyes and broader perspective would immediately identify. Internal assessors may also have to postpone or rearrange their everyday jobs/roles (e.g., security operations) to conduct internal assessments.

External assessors are those who are organizationally distinct from the assessment target. External assessors come in from the outside and review the security of the target with a more dispassionate and objective perspective. External assessments offer stakeholders an opportunity to receive outside expertise to evaluate their security and provide recommendations to their team.

> **Utility and limitations:** The primary strength of external assessments is that they offer stakeholders a chance to utilize the expertise and perspective of outsiders in assessing security. External assessors tend to be specialized in assessing security, and can offer perspectives that are informed by a much wider range of cybersecurity scenarios and actors. External assessors are also disinterested in the specifics of the target, and can provide neutral opinions detached from the *status quo* or reputations of the target's owners and operators. Of course, external assessments still have limitations. External assessors may require considerable time to learn the target's functions, mission, priorities, and nuances, which internal assessors would already have. External assessments necessarily entail information security risk, as outsiders must be trusted with sensitive information. Finally, external assessments are not completely immune to biases, particularly if the external assessor has a financial interest in recommendations made in the deliverable, such as future security assessments or purchasing security services.

---

[21] *See, e.g.,* Cyber Security Evaluation Tool, ICS-CERT, *available at* https://ics-cert.us-cert.gov/sites/default/files/FactSheets/ICS-CERT_FactSheet_CSET_S508C.pdf; "FFIEC Cybersecurity Assessment Tool," FFIEC, *available at* https://www.ffiec.gov/cyberassessmenttool.htm.

[22] *See, e.g.*, *supra* note 20.

[23] By overhead, we are referring to the upfront costs associated with familiarizing an assessor with the target, settling on processes for communication and feedback, and other internal work environment challenges that internal assessors will already be familiar with.

3.2.4 "Red" and "Blue" Assessor Roles

In real world scenarios, cybersecurity practice takes place in the context of adversarial threats. As such, an important procedural parameter is the degree to which the assessor takes on an adversarial role to the target. "Red" assessments are those where the assessor takes on the role of an adversary, approaching the target as adversary would, and often attempting to compromise the target. "Blue" assessments, by contrast, are those where the assessor coordinates and cooperates with the target in conducting the assessment, with no adversarial activities.[24] Some assessments operate at some shade of purple, where red and blue assessment methodologies are blended to optimize the assessment process.[25]

"Red" assessors are those who adopt an adversarial role to the assessment target, identifying vulnerabilities and attack vectors as an adversary would.[26] Red assessors actively engage the target system as a hostile actor, often limiting themselves to the information likely available to a given type of attacker. "Red" assessments view the target system from an adversary's eyes, and may specialize in mimicking specific threat actors, such as nation states.[27] "Red teams" have become a mainstay of cybersecurity[28], in part because of the prevalence of ex-hackers in the security community, and in part because they facilitate realism in the assessment process.[29]

> **Utility and limitations:** The primary strength of red assessors is that they bring an adversarial perspective to a realm where adversarial relationships are the driving concern. Successful red assessments are immediately actionable, as they identify specific vulnerabilities (e.g., in policy, processes, technologies) that an attacker would be able to exploit. Successful red assessments also motivate stakeholders, as the salience of a successful penetration test is similar to that of an actual breach. But more fundamentally, red assessments facilitate depth in the security assessment, identifying specific, but consequential vulnerabilities that blue teams might never have found. Finally, a successful penetration test is a useful data point, as it definitively proves the existence of a vulnerability. Alternatively, red team assessments are can sacrifice breadth for depth, providing narrower perspectives, and otherwise proving very expensive. Red team assessments also tend to be highly assessor-dependent, relying on the skill of the specific assessor and being limited by their biases, techniques, and habits. In general, these traits make red teams less consistent, harder to contextualize, and more easily misleading when taken out of context, as a successful penetration test does not necessarily equate to unacceptable risk. Perhaps most troublingly, red assessments coupled with live settings can cause real damage, breaking critical systems or seriously impacting performance.

"Blue" assessor roles support active coordination and cooperation with the target, and generally take the perspective of the defender. At the extreme, blue assessors emphasize breadth over depth, take assessment

---

[24] Note that the terms "red team" and "blue team" derive from adversarial security exercises where the red team plays the role of attacker and the blue team plays the role of defender. Here, we use "blue" more broadly.
[25] *See, e.g.,* Marszalik, P., "Purple Teaming: a Cybersecurity Assessment," Crowe Horwath, 19 Jan. 2017, *available at* https://www.crowehorwath.com/cybersecurity-watch/purple-teaming/.
[26] *See, e.g.,* "Red Team Operations," FireEye, *available at* https://www.fireeye.com/services/red-team-operations.html; "Cybersecurity Assurance Program (CAP) Red Team," NIST, May 2012, *available at* https://csrc.nist.gov/CSRC/media/Events/ISPAB-MAY-JUNE-2012-MEETING/documents/may31_cap-red-team-brief_rkaras.pdf.
[27] *See, e.g.,* IDART, Sandia National Labs.
[28] *See, e.g.,* CIS Critical Security Control 20, v6.1.
[29] *Supra*, Section 4.3.2.

targets claims at face value, and do not exhaustively double check and verify claims the target makes. As such, these assessments tend to be more programmatic, holistic, and process-focused, and pair nicely with broad assessment target definition.

> **Utility and limitations:** Blue assessments primary strength is that they allow for broader, more holistic assessment processes. Blue assessments are versatile, allowing for both breadth and depth, based on the specifics of the assessment target. This versatility also lends blue assessments toward defensive innovations, as many defenses arise without regard to specific adversarial scenarios (such as effective asset inventories and multifactor authentication). Blue assessments can also be much more efficient, as cooperation between the target and the assessors allows for immediate identification of problem areas that should be prioritized. Yet deliverables from blue assessors can be more difficult to evaluate than red teams, as they lack the immediate binary of a successful/unsuccessful compromise, and can be heavily reliant on cooperation and honesty for success. Blue teams also have a greater potential for bias, as they lack the grounding of a red assessment. Finally, blue assessors' prioritization of breadth can make them less powerful when assessing depth, missing vulnerabilities arising from bugs that a red assessor would have anticipated.

3.2.5 Quantitative and Qualitative Analyses

Cybersecurity assessments can use both quantitative and qualitative analyses to characterize phenomena observed in the world. Quantitative analysis refers to analytical processes which rely on information or data relating to quantities.[30] Qualitative analysis, by contrast, refers to analytical processes that utilize information relating to qualities that are descriptive, subjective, or difficult to measure.[31] The distinction between quantitative and qualitative analyses is well known to students of science and engineering, but can be difficult to unravel in cybersecurity.[32] True quantitative analyses in cybersecurity are few and far between, and qualitative analyses are often expressed in numerical forms, creating so-called "quasi-quantitative analysis."[33] Understanding how to identify these different methodologies and when they are useful is critical for any organization considering cybersecurity assessments.

Quantitative analyses are used to understand the world from a dispassionate, objective viewpoint, and limit the role of human interpretation in their collection and analysis. Examples of quantitative measurements include: the number of full time employees with cybersecurity duties, cybersecurity budget as a percentage of IT budget,[34] the average time from breach to response, or other metrics that rely on numerical measurement and do not allow room for human interpretation in their collection and interpretation. These objective

---

[30] *See, e.g.,* Borg, S., "Implementing a Quantitative Risk-Based Approach to Cybersecurity," RSA, Feb. 2014, *available at* https://www.rsaconference.com/writable/presentations/file_upload/str-w01-implementing-a-quantitative-approach_v2.pdf.

[31] *See, e.g.,* Cherdantseva et al., "A review of cyber risk assessment methods for SCADA systems," Computer and Security, Feb. 2016, *available at* https://www.sciencedirect.com/science/article/pii/S0167404815001388.

[32] Note also that "mixed methods" research, which combines quantitative and qualitative analyses, is commonly utilized in research communities to balance the utilities and limitations of both approaches.

[33] *See, e.g.,* Nagpaul, P., "Quasi-quantitative measures of research performance: An assessment of construct validity and reliability," Scientometrics, June 1995, *available at* https://link.springer.com/article/10.1007/BF02020567, (discussing the concept of quasi-quantitative analyses.)

[34] *See, e.g.,* Russell, Jackson, & Cowles, "Cybersecurity Budgeting: A Survey of Benchmarking Research and Recommendations to Organizations," CACR, 10 June 2016, presented at the NSF Cybersecurity Summit, 17 Aug. 2016, *available at* https://drive.google.com/file/d/0ByxarFTCEi39UnZhVURBUDlYTWc/view.

qualities make quantitative analyses often used to support standardization. Note, however, that some seemingly quantitative measurements are in fact "quasi-quantitative," as the numerical values they measure ultimately derive from qualitative assessments, such as expert scoring.[35]

> **Utility and limitations:** Quantitative analyses minimize human discretion in the data collection and analysis process, making them more objective, more consistent, more easily validated, and less susceptible to human biases. Quantitative analyses are also assessor-neutral, and allow for apples-to-apples comparisons across a wide spectrum of organizations. Moreover, quantitative analyses are more amenable to mathematical and statistical analysis, allowing for more sophisticated manipulation of the underlying data. The primary limitation of quantitative analyses is limited applicability, as the control required to produce quantitatively measurement can mask the complexity of the real world. This trait makes quantitative analyses potentially misleading, particularly when taken out of context, and can require specialized expertise to interpret correctly. Moreover, pure quantitative measurement still requires interpretation by humans, who will determine which quantitative measurements to ascribe value, the context in which they are presented, and what actions to take based on that information.
>
> Finally, it is worth noting that cybersecurity has struggled to identify meaningful quantitative measurements. Cybersecurity is difficult to study purely quantitatively, as attackers change behavior in response to defender action. And more fundamentally, cybersecurity is a new discipline, frequently disrupted by technological change, making any actuarial history limited and rapidly obsolete.

Qualitative analyses are used to characterize complex and often difficult-to-quantify phenomena. Qualitative analyses embrace the utilization of human language and interpretation in the assessment process, relying on the expertise of the assessor to synthesize conclusions from a wider range of informational sources. Qualitative analyses look to *quality,* and can assess more complex characteristics, like maturity.[36] Qualitative analyses forego the strict reproducibility and narrow robustness of quantitative ones in exchange for greater flexibility, applicability, and practical utility. Qualitative analyses are also empowered to draw inferences, extensions, and logical conclusions that are not directly produced or producible by statistical, mathematical, or computational techniques.

> **Utility and limitations:** The primary strength of qualitative analysis is the ability to characterize complex phenomena. Qualitative analyses harness the expertise of the assessor. Unlike quantitative analyses, qualitative analyses are not limited by the form and availability of data, are empowered to draw conclusions from a broad range of informational sources, and are more sensitive to the broader context. This makes qualitative analysis more adaptable, more holistic, and more easily understood in the context of the mission. But the limitations flow directly from these strengths. Qualitative assessments are less consistent, less verifiable, and less repeatable than quantitative ones. By relying so heavily on the expertise of the assessor, qualitative assessments may be more susceptible to human

---

[35] For example, a pure quantitative measurement would be the time it takes to install a patch from when it is released. An example of a related quasi-quantitative measurement would be a "score" given by an assessing organization on how well the target implements patches.

[36] *See, e.g.,* Cybersecurity Capability and Maturity Model (C2M2), Department of Energy, https://www.energy.gov/oe/services/cybersecurity/cybersecurity-capability-maturity-model-c2m2-program/cybersecurity.

fallibilities, such as biases, mistakes, and simple incompetence. Put simply, qualitative assessments are harder to evaluate, making their output less reliable for stakeholders.

4. CAPPs: Applying the Parameters as a Descriptive Tool

While the parameters are useful individually to understand different facets of a cybersecurity assessment, their true strength lies in combination. By using the parameters as a tool, stakeholders are empowered to describe the full scope of cybersecurity assessments available, and evaluate individual cybersecurity assessments in a holistic manner. Below, we offer several sample Cybersecurity Assessment Parameter Profiles (CAPPs), which serve to quickly convey Substantive and Procedural parameters/values, along with a brief justification of why we believe that categorization is appropriate.[37] In addition to characterizing existing assessments, CAPPs can be used to define a particular type of assessment, identify what characteristics that type of assessment would operate with, and help identify whether that assessment exists in the market.

To validate the parameter's utility, we have applied the parameters to a select set of sample assessments and generated CAPPs for each. Showcasing the application of the parameters in this fashion will provide useful context and help clarify how the parameters can be used in concert to understand various cybersecurity assessments. The remainder of the section will walk through several cybersecurity assessments and characterize them in terms of the parameters. We selected the specific assessments to emphasize the variety of assessments available.

Within the CAPP, each parameter is represented by the following values:

1. An extreme (e.g., "Red"), representing assessments with a clear focus on that value;
2. "Balanced," representing an even split between two extremes;
3. "Flexible," representing an assessment that can change on that parameter; or
4. "[No determination]," if we were unable to determine based on the information available.

4.1 Information Design Assurance Red Team (IDART)

<u>Prominent Parameter(s)</u>: Red, Specialized

<u>Description</u>: Developed in 1996 by Sandia, IDART emphasizes adversary-based threat modeling to identify vulnerabilities. Sandia describes IDART as utilizing "a multi-disciplinary assessment team to improve the security of critical systems through systematic analysis from an adversary perspective." IDART conducts vulnerability assessments and design assurance on a wide range of targets, ranging from individual components to enterprise organizations. IDART uses the wide-ranging expertise of Sandia's personnel to create "characterization views" that provide adversarial perspectives that model specific attacker behaviors against the security of the target. These adversarial perspectives are then prioritized based on the specifics of the adversary modeled. IDART also offers optional "engagements" which include live site visits and in-

[37] Note that our review of the assessments outlined in each CAPP is not intended to be definitive. Each CAPP was built on limited evidentiary sources. None of the characterizations below are intended to represent criticisms of the assessment at issue.

person testing. IDART produces a report that outlines its findings, emphasizing high-impact "nightmare consequences," and providing recommendations in the form of "prioritized mitigation strategies."

Sources: We based our analysis on the publicly facing documentation appearing on the IDART website, http://www.idart.sandia.gov/, which provides an overview of the process and asserts its strengths.

Parameter	Focus	Discussion
Advisory vs. Informational Deliverables	Balanced: this assessment is balanced between advisory and informational deliverables.	IDART deliverables claim to: "Identify nightmare consequences; Characterize target systems; Identify potential vulnerabilities whose exploitation will result in nightmare consequences; and Provide **prioritized mitigation strategies** so owners can make informed choices."
Broad vs. Narrow Targets	Flexible: this assessment can scope to both narrow and broad assessment target.	IDART's stated scope can include "components, devices, networks, infrastructures, and world-wide enterprises." It may be appropriate to think of IDART as a 'suite' of assessments built around Sandia's varied expertise.
Minimum, Maximum, and Tailored Standards	Tailored: this assessment tailors its standards to the assessment target.	IDART tailors its assessments based on the "requirements and expectations" of the customer, which "inform the project plan . . . specifies goals, logistics, and nature of the effort." The IDART report specifies that its deliverables are "tailored to customer's needs."
Specialized vs. Standardized Methodologies	Specialized: this assessment follows a specialized methodology.	"Sandia retains a wide range of security expertise in a variety of operational contexts that is integrated into IDART assessments to assist in the characterization and analysis of target systems." IDART is an expert-driven assessment that harnesses the wide-ranging, multidisciplinary experts who work at Sandia. Because IDART can only be performed by Sandia, and is driven by Sandia's expert personnel, we categorize it as Specialized.
Live, Conceptual, and Simulated Settings	Balanced (Conceptual/Live): this assessment uses both conceptual and live settings.	IDART emphasizes that it "often ha[s] the highest impact during the design and development phase where cooperative red team assessments cost less, and potential critical vulnerabilities can be uncovered and mitigated more easily" However, IDART does offer optional "engagements" which involve Live test settings.
Internal vs. External Assessors	External: this assessment is conducted by external assessors.	IDART can be conducted solely by a third party: Sandia National Labs.

Red vs. Blue Assessor Roles	Red: this assessment uses predominantly red assessment technique.	IDART emphasizes its use of adversary-based, "red team" techniques. Of particular note is IDART's stated capability to emulate a range of threat actors, including nation-states.
Quantitative vs. Qualitative Analyses	[No determination]	Based on available sources, we were unable to determine whether IDART focused on qualitative or quantitative analyses.

4.2 Cyber Security Evaluation Tool (CSET)

Prominent Parameter(s): Internal, Standardized

Description: CSET is a cybersecurity self-assessment tool created by the Department of Homeland Security's Industrial Control Systems Cyber Emergency Response Team (ICS-CERT). CSET was designed primarily for industrial control systems, but is applicable to any IT system. CSET is a standardized software package that utilizes a questionnaire to assess one's security posture relative to identified standards, such as those published by NIST or NERC. CSET allows for users to select specific security standards they wish to be graded against, and includes a security assurance level (SAL) to tailor the level of security the user is aiming for. CSET produces a report highlighting its findings and prioritizing recommendations based on the questionnaire, the standards selected, and the SAL.

Sources: We based our analysis on publicly facing documents on the CSET website, https://cset.inl.gov/SitePages/Home.aspx, which outlines the process and provides a link to download the tool.

Parameter	Focus	Discussion
Advisory vs. Informational Deliverables	Advisory: this assessment is predominantly advisory	CSET creates a report that "[h]ighlights vulnerabilities in the organization's systems and provides recommendations on ways to address the vulnerability" and "[t]he output from CSET is a prioritized list of recommendations for improving the cybersecurity posture of the organization's enterprise and industrial control cyber systems."
Broad vs. Narrow Targets	Flexible: this assessment can be applied to both broad and narrow targets.	CSET can be tailored to specific security documents the stakeholder wishes to grade against, e.g. NIST 800-53.CSET "includes both high-level and detailed questions related to all industrial control and IT systems."
Minimum, Maximum, and Tailored Standards	Tailored: this assessment tailors its standards to the assessment target.	CSET includes a "security assurance level (SAL)", which can be "selected or calculated and provides a recommended level of cybersecurity rigor necessary to protect against a worst-case event." Despite this extreme language, the SAL appears to operate

		primarily as a tailored, rather than maximum standard. (Note, additionally, that many of the standards CSET evaluates against operate as minimum standards.)
Specialized vs. Standardized Methodologies	Standardized: this assessment uses a standardized methodology.	CSET is a "desktop software tool" that "provides users with a systematic and repeatable approach" to cybersecurity, and "guides asset owners and operators through a step-by-step process" to securing their systems.
Live, Conceptual, and Simulated Settings	Conceptual: this assessment uses only conceptual assessment settings.	CSET generates outputs entirely based on answers to a questionnaire, and has no mechanism for validating answers or testing live system performance.
Internal vs. External	Internal: this assessment is conducted by an internal assessor.	CSET is a free software package designed to be run internally by organizations.
Red vs. Blue Assessor Roles	Blue: this assessment uses blue assessment techniques.	CSET does not identify any adversarial techniques that it utilizes.
Quantitative vs. Qualitative Analyses	Qualitative: this assessment relies predominantly on qualitative analyses.	Although CSET does not identify its analyses explicitly, the standards documents it cites and evaluates against rely predominantly on qualitative analyses.

4.3 NIST Cybersecurity Framework "Tiers"

Prominent Parameter(s): Internal, Specialized

Description: The NIST Cybersecurity Framework v1 (NIST CSF) is a collaborative effort between industry and government to create a voluntary set of resources, standards, and practices to manage cybersecurity risk. Within NIST CSF are the "Framework Implementation Tiers," a construct that allows organizations to self-assess their current state and desired state of cybersecurity by following the "Tier selection process." The Framework explicitly denies that the Framework Tiers represent a maturity model, although structurally the Framework Tiers operate in a similar hierarchical manner. The relationship between the Framework Tiers and the Framework core is unclear.

> "Tiers describe the degree to which an organization's cybersecurity risk management practices exhibit the characteristics defined in the Framework (e.g., risk and threat aware, repeatable, and adaptive). The Tiers characterize an organization's practices over a range, from Partial (Tier 1) to Adaptive (Tier 4). These Tiers reflect a progression from informal, reactive responses to approaches that are agile and risk-informed. During the Tier selection process, an organization should consider its current risk management practices, threat environment, legal and regulatory requirements, business/mission objectives, and organizational constraints."

Each of the 4 Tiers has a prose definition broken down by "Risk Management Process," "Integrated Risk Management Program," and "External Participation."

> "The Tier selection process considers an organization's current risk management practices, threat environment, legal and regulatory requirements, business/mission objectives, and organizational constraints. Organizations should determine the desired Tier, ensuring that the selected level meets the organizational goals, is feasible to implement, and reduces cybersecurity risk to critical assets and resources to levels acceptable to the organization. Organizations should consider leveraging external guidance obtained from Federal government departments and agencies, Information Sharing and Analysis Centers (ISACs), existing maturity models, or other sources to assist in determining their desired tier."

Sources: https://www.nist.gov/sites/default/files/documents/cyberframework/cybersecurity-framework-021214.pdf

Parameter	Focus	Discussion
Advisory vs. Informational Deliverables	Advisory: the assessment is focused on providing recommendations.	The primary purpose of the tier selection process is to identify gaps between the target's current state and their desired state, and recommending action to the meet the definition set out for their desired state.
Broad vs. Broad Targets	Broad: this assessment focuses on broad assessment targets.	The Tier selection process is intended to be used for "organizational cybersecurity risk management practices."
Minimum, Maximum, and Tailored Standards	Tailored: this assessment tailors its standards to the assessment target.	The Tier selection process is an internal, self-driven selection process that allows for tailoring to the target's needs. Organizations have freedom to select between the various Tiers to suit their needs. Although Tier 1 may appear to operate as an informal minimum standard, the Framework notes that "organizations identified as Tier 1 (partial) are **encouraged** to consider moving toward Tier 2 or greater, Tiers do not represent maturity levels."
Specialized vs. Standardized Methodologies	Specialized: the assessment follows a specialized methodology.	Although the Framework Tiers include descriptive language of target states, and the Framework Core references other documents, there is little procedural detail specifying how the self-assessment should be conducted.
Live, Conceptual, and Simulated Settings	Conceptual: this assessment uses only conceptual settings.	The Tier selection and self-assessment processes do not specify any simulated or live settings, although an organization could choose to deploy these settings independently.

Internal vs. External	<u>Internal:</u> this assessment is intended for internal assessors.	While CSF encourages the usage of external guidance, it generally endorses self-assessment. “Organizations should determine the desired Tier, ensuring that the selected level meets the organizational goals, is feasible to implement, and reduces cybersecurity risk to critical assets and resources to levels acceptable to the organization.”
Red vs. Blue Assessor Roles	<u>Blue:</u> This assessment relies on blue assessment techniques.	The Tier selection process does not specify any adversarial processes.
Quantitative vs. Qualitative Analyses	<u>Qualitative:</u> this assessment uses predominantly qualitative analyses.	The Tier definitions offer a high-level, language-focused standard. Apart from the numerical values of the various Tiers, there is no indication that mathematical, computational, or statistical methods are necessary.

4.4 Payment Card Industry - Data Security Standard (PCI-DSS) Audit

<u>Key Attribute(s)</u>: Minimum Standard, Standardized

<u>Description:</u> PCI-DSS is a private-sector cybersecurity standard generated by the Payment Card Industry Security Standards Council (PCI-SSC). PCI-DSS compliance is mandatory for businesses that wish to accept payment from most major credit cards. PCI-DSS is a set of minimum security controls that must be in place for credit card companies to accept transactions from that business. The PCI-DSS requirements consist of 12 broad categories of security controls, each of which is further subdivided into more narrow requirements. PCI-DSS compliance is validated by a “Qualified Security Assessor” (QSA) or an “Internal Security Assessor” (ISA), both of which are certified by the PCS-SSC.

<u>Our Assessment</u>: We based our assessment on online sources from the Payment Card Industry Security Standards council https://www.pcisecuritystandards.org/, including PCI-DSS v.3.2.

Parameter	Focus	Discussion
Advisory vs. Informational Deliverables	<u>Informational:</u> this assessment focuses on providing information.	The output of a PCI-DSS audit is a “Report on Compliance,” which produces a summary of the QSA’s findings, and only reports on the existence or non-existence of required security controls and the existence of compensating controls.
Broad vs. Narrow Targets	<u>Balanced:</u> this assessment is balanced between broad and narrow assessment targets.	PCI-DSS utilizes 12 security controls which primarily focus on the security of payment systems, (e.g. “protect stored cardholder data”), but does not assess the security of the target organization more broadly. “The PCI DSS security requirements apply to all system components included in or connected to the **cardholder data environment**.” Although

		scoped to a "system," the payment systems tend to be quite large. Nevertheless, this is a subset of the overall organization, so we call this Balanced.
Minimum, Maximum, and Tailored Standards	Minimum: this assessment uses a minimum standard.	"PCI DSS provides a **baseline** of technical and operational requirements. PCI DSS applies to all entities involved in payment card processing." "PCI DSS comprises a **minimum set** of requirements for protecting account data, and may be enhanced by additional controls and practices to further mitigate risks, as well as local, regional and sector laws and regulations"
Specialized vs. Standardized Methodologies	Standardized: this assessment uses a standardized methodology.	The PCI-DSS Audit follows a consistent structure, conducted by approved QSAs, outlined in the "Report on Compliance" template https://www.pcisecuritystandards.org/documents/PCI-DSS-v3_2-ROC-Reporting-Template.pdf.
Live, Conceptual, and Simulated Settings	Conceptual: this assessment uses only conceptual assessment settings.	The PCI-DSS audit relies on interviews, document inspection, etc., to validate compliance. Individual audit instructions follow a trend of "Identify the document" or "Identify the responsible personnel."
Internal vs. External	Flexible: this assessment can be conducted by an internal or external assessor.	PCI-DSS compliance can be validated by either an external Qualified Security Assessor," (QSA) or an "Internal Security Assessor" (ISA). Both types of assessors must be certified by PCI-SSC.
Red vs. Blue Assessor Roles	Blue: this assessment uses blue assessment techniques.	The PCI-DSS audit does not employ any red team methodologies.
Quantitative vs. Qualitative Analyses	Qualitative: this assessment uses predominantly qualitative analyses.	The PCI-DSS audit outlines a process primarily informed by qualitative analyses.

4.5 MITRE Crown Jewels Analysis (CJA)

Prominent Parameter(s): Narrow, Maximum Standard

Description: MITRE Corp's Crown Jewels Analysis, more formally referred to as "Mission-Based Critical Information Technology Asset Identification" is a cybersecurity assessment process that focuses on the identification of high criticality assets, (so-called "crown jewels"). CJA is often expected to serve as a precursor to other MITRE assessment methodologies, like "Threat Assessment & Remediation Analysis," or as part of MITRE's larger "Mission Assurance Engineering" process. CJA is designed to help stakeholders evaluate how the specific failures will impact their mission, and provide a sense what failures will have the most pronounced impact.

Our Assessment: We based our assessment on material available through the CJA website: https://www.mitre.org/publications/systems-engineering-guide/enterprise-engineering/systems-engineering-for-mission-assurance/crown-jewels-analysis.

Parameter	Focus	Discussion
Advisory vs. Informational Deliverables	Informational: this assessment focuses on providing information.	The output of CJA is the identification of high criticality assets, displaying mission impact analyses for "provides insight into ops impact of failures," helps "program officers gain understanding of user needs and system requirements," and the results "provide specific mission impacts for use in risk assessments."
Broad vs. Narrow Targets	Broad: this assessment focuses on a narrow assessment target.	CJA focuses on the identification of critical assets. CJA looks broadly, identifying critical assets across the scope of the organization.
Minimum, Maximum, and Tailored Standards	Maximum: this assessment focuses on assessing against a maximum level of security.	CJA is targeted at defending against high-level threats, such as "Advanced Cyber Threats," and assesses the threat toward crown jewels assuming these highest threat capabilities.
Specialized vs. Standardized Methodologies	Balanced: this assessment is partly standardized and partly specialized.	CJA is part of the broader Mission Assurance Engineering process, which is a "common, repeatable risk management process that is part of building secure and resilient systems."
Live, Conceptual, and Simulated Settings	Conceptual: the assessment uses predominantly conceptual assessment settings.	CJA "begin[s] during systems development and continue[s] through system deployment," and emphasizes that "identifying key system accounts, critical files, and other critical assets will require technical insights from the development team."
Internal vs. External	External: this assessment is conducted by an external assessor.	The methodology is conducted by MITRE. (The documentation suggests, but does not state, that this assessment may be able to be conducted internally as well. However, we could not make a definitive determination on this point.)
Red vs. Blue Assessor Roles	Blue: this assessment uses predominantly blue assessment techniques.	CJA relies primarily on SME questionnaires, Failure Mode-like analysis, and dependency trees. CJA does not list any primarily red-oriented assessment techniques.
Quantitative vs. Qualitative Analyses	Qualitative: this assessment uses predominantly qualitative analyses.	CJA uses "facilitated discussions with system subject matter experts" to produce "qualitatively expressed" dependencies, while using "provisions . . . to reduce subjectivity."

5. Conclusion and Use Cases

The growing variety of cybersecurity assessments and the lack of a consistent language make it difficult to clearly identify what a particular assessment does, when it will be useful, and whether it is appropriate for a particular mission. The parameters and CAPP approach presented here offer a simple conceptual framework to support informed decision-making regarding cybersecurity assessments. They offer a single, comprehensive approach to understanding cybersecurity assessments. By utilizing this simple conceptual framework, stakeholders and assessors are empowered to characterize the full scope of cybersecurity assessments, identify the utility and limitations of those different kinds of assessments, and frame discussions on the nature of an assessment's methodology and deliverables.

We propose the following use cases for the parameters and CAPP tool.

1. <u>Understanding and selecting among competing cybersecurity assessments:</u> The most natural use of the parameters is to inform stakeholders when choosing among a selection of cybersecurity assessments. By framing each of the available assessments in terms of the parameters, the stakeholder will be able to immediately identify likely strengths and limitations of each, and determine which assessment(s) are best situated to address the particular needs of their mission.

2. <u>Building a portfolio of cybersecurity assessments:</u> A closely related use of the parameters is to help stakeholders select a "suite" of assessments to meet the full needs of their mission. Although any single assessment is unlikely to address the full scope of one's cybersecurity needs, a combination of assessments (with differing parameters profiles) can be used in tandem to do what a single assessment cannot.

3. <u>Structuring conversations between stakeholders and assessors:</u> The parameters can be used as a basic lexicon to structure conversations between stakeholders (who are looking for a cybersecurity assessment) and the assessors. Although the stakeholders may not have an intimate knowledge of cybersecurity minutia, the parameters provide a basic framework when approaching conversations with assessors. The parameters help stakeholders articulate what they expect from a given assessment, provide them a set of high-level questions to determine what a specific assessment does, and give assessors a tool to contextualize their specific assessments in more general terms.

4. <u>Helping cybersecurity assessors define their own assessments:</u> The parameters can be used by the developers, owners, and operators of cybersecurity assessments as a means to convey what a particular assessment does. The parameters allow for individual cybersecurity assessments to immediately convey what they do, and help frame more in-depth discussions of individual assessments.

5. <u>Identifying gaps in currently available assessments, and developing new assessments:</u> In addition to characterizing existing assessments, the parameters can be used to identify current gaps in the assessment landscape. By exploring unusual combinations of parameters, assessors may be able to develop new or uncommon assessments with unique benefits to the broader community.

6. <u>Expanding the literature on assessments more generally, and facilitating future work comparing assessments between fields:</u> The parameters can also be used as a jumping-off point for the comparison of cybersecurity assessments and assessments in other fields and in other contexts. Although cybersecurity assessments have attributes that are unlikely to found in other contexts, (for example, education), the parameters will help frame the discussion for learning from other fields.

Appendix A: Operational Definition Analysis for "Cybersecurity Assessment"

In Section 2, we introduced our operational definition of cybersecurity assessment. In the table below, we break down and explain the grounding and meaning for each element of this definition.

> A cybersecurity assessment is an analytical activity directed at an identified target, whose outputs' purpose is to inform stakeholder cybersecurity decisionmaking regarding: (a) the characteristics and appropriate operational roles of the target; (b) the target's readiness to operate; and/or (c) resources, policy, processes, and controls warranted to support the target's operational use.

"A cybersecurity assessment…"	The US Navy defines cybersecurity broadly as: "...the prevention of damage to, protection of, and restoration of computers, electronic communications systems, electronic communications services, wire communication, and electronic communication, including information contained therein, to ensure its availability, integrity, authentication, confidentiality, and nonrepudiation"[38] The Oxford English Dictionary (online) defines assessment in the context of the verb "assess": Assessment: "The action of assessing someone or something." Assess: "Evaluate or estimate the nature, ability, or quality of"
"...an analytical activity..."	"analytical" emphasizes that cybersecurity assessments involve analysis, i.e., "a detailed examination or study of something so as to determine its nature, structure, or essential features." [OED] "activity" acknowledges that some assessments of interest may not have a discernible or repeatable process, and may simply be an "activity." An example would be an activity that meets all other elements of the definition, but emerges from assessor activity and may not have a recorded or formal process.

[38] National Security Presidential Directive 54/Homeland Security Policy 23, "Cybersecurity Policy," White House, 8 Jan. 2008, https://fas.org/irp/offdocs/nspd/nspd-54.pdf.

"...directed toward an identified target..."	Cybersecurity assessments can be directed toward organizational cybersecurity programs or the cybersecurity programs, plans, or activities of specific missions, platforms, systems of systems, systems, or system components (including hardware, software, or human elements). As such, we adopt the broad term "target."
"...whose outputs'..."	If there is no output, there is no assessment for our purposes. If a tree falls in the forest, and no one hears it, we don't care. However, assessments with a wide range of deliverables are within scope: These deliverables include, but are not limited to, written reports, briefings, conversations, and raw data.
"...purpose is to inform stakeholder decisionmaking..."	"Purpose" limits the definition to scenarios where there is an intentional relationship between assessor, assessment, output, and stakeholder. Stakeholders can include entities and individuals who are organizationally or operationally "close" to the target (e.g., the owner or operator of a target system), or those relatively distant (e.g., an oversight group; strategic leadership). Decisionmaking includes a broad range of strategic and tactical decisions (e.g., resource allocation; risk acceptance/mitigation/avoidance/transfer; control selection; incident response actions).
regarding (a) the characteristics and appropriate operational roles of the target	Intentionally very inclusive. Cybersecurity assessments can provide value by educating stakeholders on the relevant facts of the target, especially with regard to whether, when, how, why, by whom, or to what extent the target should utilized.
regarding … (b) the target's readiness to operate …	Emphasizes cybersecurity assessments can impact decisions regarding a target's readiness. Cybersecurity assessments can be conducted against conceptual designs, prototypes, units in training, and other targets not yet or not currently deployed.
regarding … (c) resources, policy, processes, and controls warranted to support the target's operational use.	Emphasizes cybersecurity assessments can impact decisions not only about whether and how to utilize the target in operations, but how to structure the operational context to support the target.

Appendix B: Our Methodology

Although prior work has begun to characterize and categorize cybersecurity assessments, particularly risk assessments,[39] we found no prior work that set out a comprehensive framework for characterizing cybersecurity assessments.

Thesis. Based on prior experience designing and conducting a range of cybersecurity assessments, we hypothesized that we could describe *any* cybersecurity assessment in terms of a set of descriptive parameters. These parameters would operate on a spectrum; each offering a range of values that are logically grouped together. For example, "red team" assessments and "blue team" assessments are common concepts in the cybersecurity community, used to describe and distinguish assessment methods. We observed distinctions like red/blue are useful not only in understanding the assessment methodology, but also in roughly gauging the nature of an assessment's utility and limitations. Both red and blue assessments are good for some purposes, but not effective for others.[40] Knowing this, we set out to identify any other descriptive parameters that could help us categorize and characterize the range of existing cybersecurity assessments.

Initial Search and Review. To uncover these parameters, we engaged in both top-down and bottom-up analysis. For our initial bottom-up analysis, we set out to review and understand a variety of cybersecurity assessments (targeted at systems, systems of systems, platforms, commands, or missions) existing and available to the Navy. These include assessments available for public consumption and ones developed and implemented in niches of the Navy and DoD community. We provide in-depth analysis of a select subset of these assessments in Section 5.

Requirements Production: Based on this preliminary review, we began articulating observed parameters, and refining our requirements for the final set. These requirements follow:

a) Each parameter should describe characteristics of cybersecurity assessments where meaningful decisions can be made regarding the design or utilization of an assessment.
b) Each parameter should help decision makers understand the utility and limitations of specific assessments. (We give particularly emphasis to characteristics whose definition, utility, and limitations may be non-obvious to stakeholders.)
c) Each parameter should describe characteristics of cybersecurity assessments where all values on that parameter have utility and limitations. This requirement excludes parameters where one value is obviously superior. Values at either extreme of a given parameter, (or somewhere in the middle) should have scenarios where that value is not just acceptable, but desirable.
d) Used as a whole, the parameters should provide practical and straightforward descriptive profiles for a range of real-world sample assessments.

Iterative Validation and Refinement: Armed with these requirements, we returned to our initial review of example assessments and analyzed the fit between our initial articulation of the parameters and these requirements. As we refined the parameter set described in Section 4, we returned to our sample set of cyber assessments to test whether our conceptual work increased alignment with real world assessments or required

---

39 For a particularly notable prior work, *see* Nachtigal, supra note 4.

40 We discuss the red and blue parameters in greater detail in Section 4.3.4.

re-grounding in reality. Our dual goals were to maximize the factual fit to the actual assessments in evidence and the functional utility of the framework.

In-Depth Application: With this more refined set of parameters in place, we began applying the parameters to a select set of specific assessments in much greater detail, looking for conceptual gaps, inconsistencies, or problems that arose.[41] Based on this analysis, we further refined the conceptual boundaries of each parameter, and worked to clarify the language describing each. We also used this phase to share the parameters with colleagues, including several cybersecurity subject matter experts (SMEs) and one educational assessment SME, to gauge clarity, completeness, and potential utility.

NOTE - Core Assumptions: Our core assumption in conducting this research is that each assessment parameter is described assuming that all other relevant variables are held equal. A number of outside factors can play a significant role in how a given assessment performs, ranging from the competence of the assessors to the money spent on a given assessment. Comparing an expertly conducted "red" assessment against an incompetently conducted "blue" assessment does not provide any true understanding of the *potential* utility and limitations of either. Conceptual excellence can always be ruined by incompetent hands; and skilled hands can often perform well with unwieldy tools. Rather than delve into these difficult-to-quantify variables, we evaluate each parameter under the assumption that these variables are held equal. In short, our goal is not to evaluate the competency of the purveyors of individual assessments, but to evaluate what those assessments should do in competent hands.

[41] The results of this in-depth analysis can be seen in Section 5: Applying the Parameters as a Descriptive Framework.